Week 7: Markov Processes

Notes by: Francisco Richter

May 15, 2025

From previous class:

Definition 0.1 (Discrete-Time Markov Chain). *A* discrete-time Markov chain $\{X_n\}_{n=0}^{\infty}$ *is a sequence of random variables taking values in a (countable) state space S such that for all $n \geq 0$,*

$$\Pr(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \;=\; \Pr(X_{n+1} = j \mid X_n = i),$$

for all states $i, j \in S$.

This property implies that the future depends on the past only through the current state.

Given a finite or countably infinite state space S, we write P_{ij} for the one-step transition probability from state i to state j:

$$P_{ij} = \Pr(X_{n+1} = j \mid X_n = i).$$

Arranging P_{ij} in a matrix $P = (P_{ij})_{i,j \in S}$ gives us the *transition matrix*, which satisfies:

$$P_{ij} \geq 0, \quad \text{and} \quad \sum_{j \in S} P_{ij} = 1 \quad \text{for each } i \in S.$$

1 Multi-Step Transitions and the Chapman-Kolmogorov Equations

1.1 Definitions

Define the n-step transition probabilities as

$$P_{ij}^{(n)} \;=\; \Pr(X_{k+n} = j \mid X_k = i),$$

for any $k \geq 0$ and $i, j \in S$. In particular, $P_{ij}^{(1)} = P_{ij}$ are the one-step probabilities.

Theorem 1.1 (Chapman-Kolmogorov). *For any nonnegative integers $n, m \geq 0$,*

$$P_{ij}^{(n+m)} \;=\; \sum_{k \in S} P_{ik}^{(n)} \, P_{kj}^{(m)}.$$

In matrix form, if $P^{(n)} = (P_{ij}^{(n)})$, then

$$P^{(n+m)} \;=\; P^{(n)} \cdot P^{(m)}.$$

Hence, $P^{(n)} = P^n$ (the n-th power of the matrix P).

Example 1.1 (Weather Model: Two States)**.** *Consider a two-state weather chain where* 0 = *Rainy,* 1 = *Sunny. Suppose*

$$P = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}.$$

Then

$$P^2 = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}^2 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}, \quad P^4 = (P^2)^2 = \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}.$$

Thus, if it is currently raining (state 0), the probability of it raining again 4 days from now is $P_{00}^{(4)} \approx 0.5749$*.*

Example 1.2 (Extended Weather Model: Four States)**.** *Consider a chain that tracks the weather on two consecutive days, thus having four states:*

$$0: (\textit{Rain, Rain}), \quad 1: (\textit{No Rain, Rain}), \quad 2: (\textit{Rain, No Rain}), \quad 3: (\textit{No Rain, No Rain}).$$

If the transition matrix is

$$P = \begin{pmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{pmatrix},$$

we can compute P^2 *to get the 2-step probabilities. For instance, if the chain starts in state* 0 *(meaning the last two days were both rainy), the chance that the next two days also include at least one rainy day can be derived from particular entries of* P^2*. This example illustrates how higher-dimensional Markov chains can encode memory of previous states, albeit at the cost of an enlarged state space.*

2 Classification of States and Long-Term Behavior

Understanding how a Markov chain behaves over many steps requires classifying its states and determining whether certain long-term distributions exist.

2.1 Communicating Classes and Irreducibility

Definition 2.1 (Communicate, Class)**.** *States* i *and* j communicate *if* $P_{ij}^{(n)} > 0$ *for some* n *and* $P_{ji}^{(m)} > 0$ *for some* m*. A set of states* C *is a* communicating class *if every pair of states in* C *communicate and no state outside of* C *communicates with a state in* C*.*

Definition 2.2 (Irreducible Markov Chain)**.** *A Markov chain is* irreducible *if the entire state space* S *is one single communicating class, i.e., one can get from any state* i *to any state* j *in a finite number of steps (with positive probability).*

2.2 Recurrence and Transience

Definition 2.3 (Recurrence/Transience)**.** *A state* i *is* recurrent *if starting from* i*, the expected number of visits to* i *is infinite; equivalently, the probability of returning to* i *at some time in the future is 1. If that probability is less than 1, then* i *is* transient.

In finite Markov chains, irreducible classes are automatically recurrent (and at least one class may be absorbing if there's a state with $P_{ii} = 1$).

2.3 Periodicity

Definition 2.4 (Period)**.** *The* period *of a state i is*

$$d(i) \;=\; \gcd\{\, n \geq 1 : P_{ii}^{(n)} > 0\}.$$

If $d(i) = 1$, we say i is aperiodic. *A Markov chain is aperiodic if all its states are aperiodic. In an irreducible chain, it suffices to check just one state.*

If a Markov chain is irreducible and aperiodic (i.e., *ergodic*), then it enjoys a host of powerful limit theorems.

2.4 Stationary and Limiting Distributions

A probability vector $\pi = (\pi_1, \pi_2, \dots)$ is called a *stationary distribution* if

$$\pi P = \pi \quad \text{and} \quad \sum_{i \in S} \pi_i = 1.$$

For a finite irreducible aperiodic chain, there exists a unique stationary distribution π, and moreover,

$$\lim_{n \to \infty} P_{ij}^{(n)} = \pi_j, \quad \text{for every } i, j \in S.$$

This means the chain forgets its initial state in the long run and converges to π.

Example 2.1 (Market Chain Convergence)**.** *Consider the* 3×3 *matrix*

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \\ 0.1 & 0.7 & 0.2 \end{pmatrix}.$$

Numerical powers P^n for large n show that each row converges to the same vector

$$\pi \approx (0.3426,\; 0.3519,\; 0.3055).$$

Hence, if you track states as "Bull", "Bear", and "Stagnant" markets, in the long run, the chain spends around 34.26% *of the time in the first state,* 35.19% *in the second, and* 30.55% *in the third, irrespective of the initial condition.*

3 Absorbing Markov Chains

A Markov chain is *absorbing* if it has at least one state i with $P_{ii} = 1$ (such a state is called *absorbing*), and from every state in the chain, there is some way (positive-probability path) to eventually enter an absorbing state.

3.1 Canonical Form and Fundamental Matrix

One typically reorders the states so that absorbing states come last, yielding a transition matrix in the form

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix},$$

where Q is the transition matrix among *transient* states and I is an identity matrix for the absorbing states. The *fundamental matrix* is

$$N = (I - Q)^{-1}.$$

Its (i, j)-th entry N_{ij} is the expected number of visits to state j starting from i, before absorption occurs. The matrix NR then gives absorption probabilities into each absorbing state.

Example 3.1 (Simple Absorbing Chain).

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{pmatrix}.$$

State 1 *(the first row) is absorbing since* $P_{11} = 1$. *One can reorder states if needed to analyze how states* 2 *and* 3 *eventually get absorbed.*

4 Branching Processes

Branching processes model how populations evolve when each individual reproduces independently of others. The canonical example:

Definition 4.1 (Galton-Watson Process). *Let* $Z_0 = 1$. *Each individual in generation* n *produces a random number of offspring in generation* $n + 1$ *according to a fixed distribution* $\{P_k\}_{k=0}^{\infty}$. *Formally,*

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i},$$

where $X_{n,i}$ *are i.i.d. with* $\Pr(X_{n,i} = k) = P_k$.

One key question is whether the population eventually dies out (i.e., hits $Z_n = 0$ for some n). Define the generating function

$$f(s) = \sum_{k=0}^{\infty} P_k \, s^k.$$

Then the extinction probability π_0 is a fixed point of f, i.e., π_0 satisfies $\pi_0 = f(\pi_0)$.

Example 4.1 (Binary Offspring). *If each individual has 0 or 2 offspring with probability 0.5 each, then*

$$f(s) = 0.5\, s^0 + 0.5\, s^2 = 0.5 + 0.5\, s^2.$$

Setting $\pi_0 = f(\pi_0)$ *gives* $\pi_0 = 0.5 + 0.5\,\pi_0^2$. *One finds that* $\pi_0 = 1$ *is the relevant solution here, indicating eventual extinction with probability 1 in this* critical *case.*

5 The Gambler's Problem (Classic)

Even without actions, the classical gambler's ruin scenario can be seen as a simple Markov chain on $\{0, 1, \ldots, G\}$ with absorbing states at 0 and G. At wealth s, the gambler wins the next coin toss with probability p and moves to $s + 1$, or loses with probability $1 - p$ and moves to $s - 1$. Setting

$$P(s) = \Pr(\text{reach } G \mid \text{start at } s),$$

yields the difference equation

$$P(s) = p\, P(s+1) + (1-p)\, P(s-1),$$

with $P(0) = 0$ and $P(G) = 1$. The solution is

$$P(s) = \begin{cases} \dfrac{s}{G}, & p = 0.5, \\ \dfrac{(p/(1-p))^s - 1}{(p/(1-p))^G - 1}, & p \neq 0.5. \end{cases}$$

This fundamental example underlies many gambler-like Markov chain models.